

SUBSTITUTE SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: GENSET SA
- (B) STREET: 24 RUE ROYALE
- (C) CITY: PARIS
- (E) COUNTRY: FRANCE
- (F) POSTAL CODE: 75008

(ii) TITLE OF INVENTION: HUMAN DEFENSIN DEF-X GENE AND DNAC COMPOSITION
CONTAINING SAME AND DIAGNOSTIC AND THERAPEUTIC APPLICATIONS

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
- (B) STREET: 2421 N.W. 41st Street, Suite A-1
- (C) CITY: Gainesville
- (D) STATE: Florida
- (E) COUNTRY: USA
- (F) ZIP: 32606

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER (unassigned)
- (B) FILING DATE: OCTOBER 18, 2001

(vii) PRIORITY APPLICATION DATA:

- (A) APPLICATION NUMBER 09/486,580
- (B) FILING DATE: FEBRUARY 25, 2000

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Frank C. Eisenschenk, Ph.D.
- (B) REGISTRATION NUMBER: 45,332
- (C) REFERENCE/DOCKET NUMBER: GEN-100D1

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4415 BASE PAIRS
- (B) TYPE: NUCLEOTIDE
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(ix) FEATURE:

- (A) NAME/KEY: Exon 1
- (B) LOCATION: 1836..1874

10045180.101801

- (ix) FEATURE:
 (A) NAME/KEY: Exon 2
 (B) LOCATION: 3394..3577
- (ix) FEATURE:
 (A) NAME/KEY: Exon 3
 (B) LOCATION: 4161..4380
- (ix) FEATURE:
 (A) NAME/KEY: start CDS
 (B) LOCATION: 3406..3408
- (ix) FEATURE:
 (A) NAME/KEY: stop CDS
 (B) LOCATION: 4276..4278
- (ix) FEATURE:
 (A) NAME/KEY: polyadenylation site
 (B) LOCATION: 4374..4379
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ACACCATTGT TCTTCATGTA ACCCCATTAG CTATACCCCTC TAGTGCAAGG AAACCATAGG 60

GCCTAGGTGCA CACCATGAGG CTGCNCTTAC AAGTTATGCA AAAACTATGG ACTTGGGAGA 120

CCTGTGCGTA ACAACATCAC ACNCCAAATT TAACACAGCTC TCCCATAAC AGCACGCTCA 180

TGTGTTACTG AGGAAATGCC TGTGGATTGG AGTGTGTTCT GTGTGCAGGA GGCTGGTGCA 240

GGTTTCACCT CTGCAGGACA CTGGACGTTT CCCAAACCA GCAGACTTTC CCCACGTGCA 300

CACACACCCC TTCTCATTTT GCTCTACAT CCATATCCAC TGGGCCCTTC AGGCACCTAC 360

TAATGCCCTA GAACCTAAAA CCATCATCTG GGGCCCAATT CCCTGAATGG CCCTAATCTC 420

TTCTCTGCTG GGAATGAGTC CAGTGCCAC TTCCTCCAAC GGTGAAATTG CTGGGCTGCT 480

ACAGATCAGG AACTCACTGC TTCTCATAG GGGCAGCCGA CTCACTGCT CTGCAACAGC 540

GACCAACCCCT AGCAGGCTT GAGATGCCTC TTGCCTCCTT AAGACTGAGG GAGACGCTTC 600

AGCTCTCACT CCACTGCCCC AAGTCCTCCA CAGCGCGTG CTGCTGCCT TCACACAGAG 660

CTGCAGGGGN AGTCTCTGTG TATCCGGCTC GCTGGACCAG CGCTGTGCAC AACCTCCCA 720

TGGCAACAGT GGCTGCCCGG CCTGCACACT GGGCTTGGA ACCTCGCTGT AGGTATTTAT 780

TCCCTCAGGA GTGACTGCAT TCTTTTCCCA TTTCAGAAA ACTGATGCCA TTACCTCAC 840

TATGAGGAGG AGGAGGAGGA GGAGGTTGGA GAGTGTGACA TTTTAAATG TGCACTATTC 900

TCCCTAGGAC TCCCCTCAA ATAACCCAGG AGGGACCATA CCAGCTCATT CCTGTGTATC 960

CCAAGCATAN GAGTAATCAT CCCACTCATG CTGAGTGTAT GGTGGCCATT AAGCCTGCCC 1020

TGAACTGGCT TTAGAACAAG GTGTTTGAGC ACACAGCACC GTCTTGCTGC CACCTTGGCC 1080

CCCTCCCTTG TGAGACCTCT GAGACACATT NAGGTCTCAC CTAAAAATCT CAGGATTCTT 1140

AGGCCCAAAN CGGTCCTAAA AAATTGTTCA GTCTGAACTC TCTAAGGTCA AGAGAAGAGG 1200
 TGGTTGCTCC CTCTAAGAAA CCACATGTTG CATGTACATC CTTAATTCCG GAAAGTCCAA 1260
 CAAACCTGCC CTGCTTAGCA ACACAAGCCG AGGTGGTACT CCTCTCACC GGGCATTTCTC 1320
 CAACACACCT GTTTGTCCAA ACAGCTTTGA TTTGTTTTTA TAGTTGGACC CCAGGTTCCC 1380
 AGGAGGCTGG TTCAGGCCAT ATTCCAAATC CTCATCTGTG TGTGAGTGGC ATTTCTAGCC 1440
 TAGCCTCCTT ACAGGGTGGG TACTATGATA CACAGCCAGG CTGTCCCAGT GGCTTTCAAT 1500
 ATTCTTTTGG TCCAGATAGT TCAGCCTCAG CACCAAGTGA GGCATCACAG GGTCAATTGT 1560
 CTTAGGAGTC ATGGAGAATT CATAGTTGGT AGCTACCTGG GCCTGGCCAG GGCTGACCAT 1620
 AGACAAGGCA TCCCTCTGTG AACTCCTATT TTAATGCCAG CTTCCCAACA AATTCTCTCA 1680
 CTGCTCTTAC CAGCAGGTAT TTAAACTACT CAATAGAAAG TAACCTGAA AATTAGGACA 1740
 CCTGTTCCCA AAGACCCCTT AAATAGGGGA AGTCCTTTTC CTGCTTGTGC ACAGCTGCTG 1800
 ATGTGGCAAC ATGAGGCCCTG GGACAGGGGA CTGTCTCTG CCCACTCTGG TAGCCTCAGC 1860
 TAGCTTAACA ATCTGTCACT AATACAATAC AAAACTTAAA CTTTCATACT GCGGTTCCAC 1920
 CCAGGAAGCT GTGTTCCCAA TCTGACCCGT GATTATGGGG CCACCTCAGA GGGNACCAG 1980
 TGAGGGAATA TTTTGCCATC TGGGACTGTT GGTGCTGGG GGCAGTGGCT ATGAGCTCAG 2040
 TTAATAAACT CAAGCAGTTT CCTTCCAAAC ACACATGTCC TACTTAACGT GTCCAACAGA 2100
 GATGATCATA CTCATANGCT GCTAAAACAT TANTTTTATT TTGAGAAAAG TCTATTCTAG 2160
 TTCCTGGCCC ATGGAGTTTT CATTINATTA NTTTTATTAT TTTGCAGAGA TGGAGTCTCA 2220
 CTATGTTGCT CAAGCTGGTC TCCAACCTCT GGGCTCAAGC GATCTTCCTA CTTTGGCCTT 2280
 TGAAAGCGCT GAGATTGCCT GTGTGAGCCA TCATGGGGGC TCACTGCCCC ACTGATTAAT 2340
 CAGATTAAAT GTTTTTTGCT ATTGAANTTG TTTGACTTCC TTGTATATTC GGATATTTAC 2400
 CCATTCTAAC ACGTAGGGTT TGCAAAATATT TTCTCTCATG TTCTGTGTTG CCTTTTCACT 2460
 CAGTTGATGG TTTCCTTTGC TGTGCAGGTG CTTTAGTGTT CAACGCAGCC CCGCTGTGCT 2520
 ATTTTCCATT TTAATGCTG TCCCTTTGAT GTCATAGCCA AGAAATAAAT GCCCAGATTA 2580
 ATGTCAAAA GCTTTATCCC TATATATTCT TCTAGTAGTT TATGGTTTCA GATCTTATGT 2640
 TTAGTCTCTC AATCCATTGA GTTGATTTTT GTATGTGGTA TAAGAAAAAA GACCACATGT 2700
 ATACATATCT CAAATTCTAA GGTAGTATAT ATTAGACACA TACAATGTGT CTATTTACAC 2760
 ACATTGAGCT GAAAATAATA AACATATTTT TATCTTTCAA TCAACTCTAT CTCTATCTCA 2820
 CTGAACCTGT TTCACCTATA GCCTGATGAG GTTGCTGTCC TCTCTACCCC AGCTCCTATA 2880
 GGAGACTGCT CATCCCCTAA CCTCAAAAAC CCCTTCATGA GGGTGATAAT GCCCTGAAT 2940
 CCTGCAATGA ATTAGTTCTC TACTACAGTG GAATTCAGGT CTGTTATGAG GGTCTGGATC 3000

10045180.101801

TCTGAAGAGA AGAGCTCTCA TTTTCAGAAA ATAAGCAGGA TTTATTCCTT GAAATTACTG 3060
 AATTAAATCA CTGTTTCGAT TACTTTTTCG AATATTAATA GTAAATATTT AACACAGTAA 3120
 AAACAGAAAT AATGGTAGGG TCCTTATCAT CACCGTGAAT TCCAGCTAG CATAGACACT 3180
 AAACCTAGAG ATTCACTA GAATGAAAGC TGGGAGAGCA GAGGAGTCTC AGAAGGATGT 3240
 GGAGGCCAAT GGACACCTGC AACCTCTCCA ACGAAATGCC TACCTCTCTT CACTGCAGCA 3300
 TCCATCTCTG AGCCTTCTCG CAGCAGAGCT ATAAATTCAG CCTGGCTCCT CCGTTCACC 3360
 ACATCCACTC CTGCTCTCCC TCCTCTCTCT CAGGTGACTA CAGTTATGAG GACCTTCACC 3420
 CTCCTCTCTG CTTTCTCTCT GGTGGCCCTT CAGGCCTGGG CAGAGCCGCT CCAGGCAAGA 3480
 GCTCATGAGA TGGCAGCCCA GAAGCAGCCT CCAGCAGATG ACCAGGATGT GGTCAATTAC 3540
 TTTTCAGGAG ATGACAGCTG CTCTCTTTCAG GTTCCAGGTG AGAGATGCCA GCATGCAGAG 3600
 CTACAGACTA GACAGAAGGA CAGGAGACAG GCTCTGGAAT TGGATCTCAG TGGCAGATGT 3660
 CACTTAGGTG GCTATACTTA ACATCTCTGG TCCTGGGATT TCTCATATCT AAATGGAATA 3720
 GAGAACCAAA GAAATCTAAG AGATTTTCTT TTCTCCAAAA ACTTGATTCC AAGATATGAC 3780
 TGTGAAATTC ACTAGATTTA AGATAAAGG AGATGCTACC TAGTTCTCTC TGGAGCCAGA 3840
 CAAACAAGCT TAAGTATATA GGAAATATTT TCACCCTGTC TATATAGGAG GTTTTAGAAC 3900
 CTGGAGAGGA GCCTAAGAAT GTGTTCAAGT GTGTGTGTGA TGGGCAGGAA TGCAGAAAAA 3960
 TGAAGCAAAG GAGAATGAGT CTCGAATCCT GTGTGACCAG CACTGCTCTG TGTATTTATT 4020
 CCTATTGACT GAGATTGTTT GTGTACCCGG CTGTAATACA GCCAACATCA CTCATCAGCC 4080
 AACATGTGAC TTCTCCAAGA TTCCCTTTAC CACCCACTGC TGNACCCCGT ACTCAGTTTC 4140
 TGATGCTCTC TCTGGGTCCC CAGGCTCAAC AAAGGGCTTG ATCTGCCATT GCAGAGTACT 4200
 ATACTGCATT TTTGGAGAAC ATCTTGGTGG GACCTGCTTC ATCTTGGTG AACGCTACCC 4260
 AATCTGCTGC TACTAAGCTT GCAGACTAGA GAAAAAGAGT TCATAATTTT CTTTGAGCAT 4320
 TAAAGGGAAT TGTATTCTTT ATACCTTGTC CTCGATTTCG TGFCTCATC CCAATAAAT 4380
 ACTTGGTAAC ATGATTCCG GGTTTTTTTT TTTT 4415

(2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTIC:
 - (A) LENGTH: 453 BASE PAIRS
 - (B) TYPE: NUCLEOTIDE
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: cDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CTCTGCCAC TCTGGTAGCC TCACGTAGCT TAACAATCTG TGACTACAGT T ATG AGG 57
Met Arg
1

ACC CTC ACC CTC CTC TCT GCC TTT CTC CTG GTG GCC CTT CAG GCC TGG 105
Thr Leu Thr Leu Leu Ser Ala Phe Leu Leu Val Ala Leu Gln Ala Trp
5 10 15

GCA GAG CCG CTC CAG GCA AGA GCT CAT GAG ATG CCA GCC CAG AAG CAG 153
Ala Glu Pro Leu Gln Ala Arg Ala His Glu Met Pro Ala Gln Lys Gln
20 25 30

CCT CCA GCA GAT GAC CAG GAT GTG GTC ATT TAC TTT TCA GGA GAT GAC 201
Pro Pro Ala Asp Asp Gln Asp Val Val Ile Tyr Phe Ser Gly Asp Asp
35 40 45 50

AGC TGC TCT CTT CAG GTT CCA GGC TCA ACA AAG GGC TTG ATC TGC CAT 249
Ser Cys Ser Leu Gln Val Pro Gly Ser Thr Lys Gly Leu Ile Cys His
55 60 65

TGC AGA GTA CTA TAC TGC ATT TTT GGA GAA CAT CTT GGT GGG ACC TGC 297
Cys Arg Val Leu Tyr Cys Ile Phe Gly Glu His Leu Gly Gly Thr Cys
70 75 80

TTC ATC CTT GGT GAA CGC TAC CCA ATC TGC TGC TAC TAA GCTTGACAGAC 346
Phe Ile Leu Gly Glu Arg Tyr Pro Ile Cys Cys Tyr *
85 90 95

TAGAGAAAA GAGTTCATAA TTTTCTTTGA GCATTAAAGG GAATTGTAT TCTTATACCT 406

TGTCCTCGAT TTCTGTCCT CATCCCAAT AAATACITGG TAACATG 453

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 94 AMINO ACIDS
(B) TYPE: AMINO ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens
- (ix) FEATURE:
(A) NAME/KEY: SIGNAL PEPTIDE
(B) LOCATION: 1..19
- (ix) FEATURE:
(A) NAME/KEY: PRO REGION
(B) LOCATION: 20..63
- (ix) FEATURE:
(A) NAME/KEY: MATURE PEPTIDE

(B) LOCATION: 64..94

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```

Met Arg Thr Leu Thr Leu Leu Ser Ala Phe Leu Leu Val Ala Leu Gln
1      5      10      15
Ala Trp Ala Glu Pro Leu Gln Ala Arg Ala His Glu Met Pro Ala Gln
20      25      30
Lys Gln Pro Pro Ala Asp Asp Gln Asp Val Val Ile Tyr Phe Ser Gly
35      40      45
Asp Asp Ser Cys Ser Leu Gln Val Pro Gly Ser Thr Lys Gly Leu Ile
50      55      60
Cys His Cys Arg Val Leu Tyr Cys Ile Phe Gly Glu His Leu Gly Gly
65      70      75      80
Thr Cys Phe Ile Leu Gly Glu Arg Tyr Pro Ile Cys Cys Tyr
85      90

```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 AMINO ACIDS
- (B) TYPE: AMINO ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: SIGNAL PEPTIDE

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```

Met Arg Thr Leu Thr Leu Leu Ser Ala Phe Leu Leu Val Ala Leu Gln
1      5      10      15
Ala Trp Ala

```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 AMINO ACIDS
- (B) TYPE: AMINO ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PRO REGION

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Glu Pro Leu Gln Ala Arg Ala His Glu Met Pro Ala Gln Lys Gln Pro
 1 5 10 15
 Pro Ala Asp Asp Gln Asp Val Val Ile Tyr Phe Ser Gly Asp Asp Ser
 20 25 30
 Cys Ser Leu Gln Val Pro Gly Ser Thr Lys Gly Leu
 35 40

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 AMINO ACIDS
- (B) TYPE: AMINO ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: MATURE PEPTIDE

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Ile Cys His Cys Arg Val Leu Tyr Cys Ile Phe Gly Glu His Leu Gly
 1 5 10 15
 Gly Thr Cys Phe Ile Leu Gly Glu Arg Tyr Pro Ile Cys Cys Tyr
 20 25 30

10045180.101801